A NOTE ON THE BAYESIAN APPROACH TO VARIANCE COMPONENT ESTIMATION FOR UNBALANCED DATA

Kathryn Chaloner

University of Minnesota
School of Statistics

Technical Report #454

December 1985

University of Minnesota
School of Statistics
Department of Applied Statistics
St. Paul, Minnesota 55108

The estimation of variance components in experiments with unbalanced sample sizes in studied. A simulation study is presented which indicates that modes of posterior distributions have good sampling properties compared to other estimators. The posterior distributions are calculated using a non-informative prior distribution which is uniform on the intra-class correlation. A simulation study for the estimation of the ratio of the variance components is also presented as is a study of the sampling properties of highest posterior density regions for this ratio. Bayesian estimators appear to be viable competitors to the many classical alternatives in a sampling framework.

KEY WORDS: Variance component analysis; Bayesian Inference; maximum likelihood; simulation.

1. Introduction

Models with components of variance are frequently used in quality control, animal breeding experiments and elsewhere. For balanced data variances are usually estimated using the minimum variance unbiased estimators (MVUE's) based on the sums of squares appearing in the analysis of variance table. For unbalanced data no MVUE's exist and the sums of squares from the ANOVA table are not sufficient statistics. Various estimators have been proposed. For example, maximum likelihood estimators (MLE), minimum norm quadratic unbiased estimators (MIVQUE), and several more variations of these approaches [see e.g. Searle (1979)]. Another common approach is to use analysis of variance (ANOVA) estimators, which are obtained by equating mean squares to their expected values. With the exception of the ANOVA estimators, all approaches for unbalanced data require substantial computational effort.

A recent paper by Swallow and Monahan (1984) included an extensive simulation study of estimators of variance components. They studied a one-way random effects model for several patterns of unbalanced data. They found average squared errors and biases for five estimators. Their results indicate that by these criteria the ANOVA estimators and the MLE's are generally preferable. No Bayesian estimators were included in the study. Their study is reproduced here, but comparing only three estimators: ANOVA, MLE's and the Bayesian estimator given by the mode of the joint posterior distribution of the variance components. A non-informative prior distribution is used. The simulation results indicate that the estimators derived from the posterior mode do have good sampling properties and are generally superior to other estimators in terms of average squared error.

In a strictly Bayesian framework of inference the sampling distribution of aspects such as means or modes of posterior distributions are not of interest (see e.g. Box and Tiao (1973), page 310). As it appears that posterior modes do seem efficient in terms of mean squared error, it would seem that they should be seriously considered in either a Bayesian or a frequentist framework.

Philosophical arguments aside, using a Bayesian approach to variance component estimation has several other practical advantages over a classical approach. First, we never find ourselves reporting a negative estimate for a variance, and an interval estimate such as a highest posterior density region will not include negative values. (This is a contrast to the ANOVA estimators, for example, where the sampling distributions are such that, for common values of the variances, negative estimates occur with quite high probability.) Second, we can report the whole of our posterior probability distribution, not just a single number, and report some measure of posterior precision. Classical estimators generally have intractable sampling distributions and standard errors are hard to calculate.

The Bayesian approach to variance component estimation has been studied extensively for balanced data. Tiao and Tan (1965) and Box and Tiao (1973) successfully implemented a Bayesian approach for balanced data. They derived some closed form estimators for the components of variance and some approximations for their posterior distributions. Another approach is used in Hill (1977). Klotz, Milton and Zacks (1969) investigated the sampling properties of Bayesian and other estimators. Skene (1983) gives computational methods for finding marginal posterior distributions for variance components of crossed and nested models. All of this work is for the balanced case.

In a Bayesian context there is little difference between the balanced and unbalanced case. Posterior distributions for variance components are derived in Hill (1965). All of the necessary intergrations are given and discussed by Hill. These posterior distributions are reviewed in section 2. The choice of a prior distribution is discussed in section 3 and a convenient non-informative prior distribution is suggested and motivated in section 3. The computational methods for implementing the Bayesian approach are outlined in section 4, and the simulation results are presented in section 5. Also in section 5.2 there is a brief investigation of the impact of using an informative prior distribution for the between group variance. The informative prior distributions are vague in that they have a mode but have infinite moments. The estimation of the ratio of the two variances is discussed in section 5.3. A small simulation study is described which examines highest posterior density regions for this ratio.

2. Posterior Distributions

We will assume that we have the usual random effects one-way analysis of variance model. There are I groups and n_i observations in each group, y_{ij} being the j-th observation in the i-th group. The group effects are independent, normally distributed and have variance σ_B^2, and the measurement errors are independent, normally distributed and have variance σ_W^2. The overall mean is μ. Thus if $e_i \sim N(0,\sigma_B^2)$ $i=1,\ldots,I$ and $e_{ij} \sim N(0,\sigma_W^2)$ $j=1,\ldots,n_i$ we have

$$y_{ij} = \mu + e_i + e_{ij} \quad . \tag{2.1}$$

In a quality control situation the groups might be batches of a chemical product and n_i samples are taken from batch i and measurements made. In an animal breeding experiment the groups might be individual cows, and a measurement is made on each of the n_i calves of the i-th cow. In a well designed experiment the n_i's are usually set to be equal. In some situations, however, the n_i's are not all equal and we have unbalanced data to analyze. This may be due, for example, to not all cows having the same number of calves, to spoilt samples and missing data, to different batches having different sampling costs, or to perhaps badly designed experiments. Interest is in the estimation of σ_B^2, σ_W^2 and possibly μ.

The model (2.1) has also been used in the Bayesian hierarchical framework of Lindley and Smith (1972) for both unbalanced and balanced data. In that situation, however, estimation of the group effects, the e_i's, is of primary interest. We will not consider that situation here.

For unbalanced data the group means, $\bar{y}_{i.}$, $i=1,\ldots,I$ and the within group sum of squares

$$s_W^2 = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2$$

are sufficient for $(\mu,\sigma_B^2,\sigma_W^2)$. The data has a multivariate normal distribution and if we let

$$\sum_{i=1}^{I} n_i = N$$

the likelihood $p(\underline{y}|\mu,\sigma_B^2,\sigma_W^2)$ is

$$p(\underline{y}|\mu,\sigma_B^2,\sigma_W^2)$$

$$\propto (\sigma_W^2)^{-(N-I)/2}\left(\prod_{i=1}^{I} (\sigma_W^2 + n_i\sigma_B^2)^{-1/2}\right)$$

$$\cdot \exp\left[-\frac{1}{2\sigma_W^2}\left(s_W^2 + \sum_{i=1}^{I} \frac{n_i}{(1+n_i\sigma_B^2/\sigma_W^2)}(\bar{y}_{i.}-\mu)^2\right)\right] \cdot \qquad (2.2)$$

To compute the posterior distribution we must choose a prior distribution $p(\mu,\sigma_B^2,\sigma_W^2)$. There is no convenient choice of prior distribution. We can make the not unreasonable assumption that a priori scale and location are independent and $p(\mu,\sigma_B^2,\sigma_W^2) = p(\mu)p(\sigma_B^2,\sigma_W^2)$. A reasonable choice of $p(\mu)$ might be uniform.

The non-informative choice of $p(\sigma_B^2,\sigma_W^2)$ is not so straightforward and we consider several choices of $p(\sigma_B^2,\sigma_W^2)$ in the next section.

Leaving our choice of prior distribution as initially $p(\mu,\sigma_B^2,\sigma_W^2) = p(\sigma_B^2,\sigma_W^2)$, (zero unless $\sigma_B^2,\sigma_W^2 \geq 0$) the joint posterior distribution $p(\mu,\sigma_B^2,\sigma_W^2|\underline{y})$ can be obtained by multiplying the likelihood (2.2) by the prior distribution and normalizing . Integrating over μ leads to an expression for $p(\sigma_B^2,\sigma_W^2|\underline{y})$, the joint posterior distribution of σ_B^2 and σ_W^2. Details of this derivation are given in Hill (1965). We define, for i=1,...,I,

$$w_i = n_i(1 + n_i\ \sigma_B^2/\sigma_W^2)^{-1} \text{ and } \hat{\mu} = \sum_{i=1}^{I} w_i\bar{y}_{i.} / \sum_{i=1}^{I} w_i \tag{2.3}$$

and then for $\sigma_B^2,\ \sigma_W^2 \geq 0$

$$p(\sigma_B^2,\sigma_W^2|\underline{y}) \propto p(\sigma_B^2,\sigma_W^2)(\sigma_W^2)^{-(N-1)/2}[\prod_{i=1}^{I}(1+n_i\sigma_B^2/\sigma_W^2)^{-1/2}]$$

$$\cdot [\sum w_i]^{-1/2}\exp\{-\frac{1}{2\sigma}[s_W^2 + \sum^{I} w_i(\bar{y}_{i.} - \hat{\mu})^2]\}\ . \tag{2.4}$$

We will study aspects of this posterior distribution for several choices of $p(\sigma_B^2,\sigma_W^2)$. These choices are all of the form (2.5) below which leads to some simplification. We let f and g be suitable functions with a suitable choice of λ and let the prior distribution be:

$$p(\mu,\sigma_B^2,\sigma_W^2) \propto (\sigma_W^2)^{-1-\lambda/2}f(\frac{\sigma_B^2}{\sigma_W^2})\cdot\exp\{\frac{-1}{2\sigma_W^2}\ g(\frac{\sigma_B^2}{\sigma_W^2})\}\ . \tag{2.5}$$

This form of prior distribution leads to simplification in the posterior distribution. We define $\tau^2 = \sigma_B^2/\sigma_W^2$ and then change variables from $(\mu,\sigma_B^2,\sigma_W^2)$ to (μ,τ^2,σ_W^2). Then for a prior distribution of the form (2.5) the posterior distribution of (τ^2,σ_W^2) can be written as in (2.6) below.

$$p(\tau^2,\sigma_W^2\,|y)$$

$$\alpha(\sigma_W^2)^{-(N+1+\lambda)/2-1}\, f(\tau^2)\Big[\prod_{i=1}^{I}(1+n_i\tau^2)^{-1/2}\Big]$$

$$\Big[\sum_{i=1}^{I} w_i\Big]^{-1/2}\exp\Big\{-\frac{1}{2\sigma_W^2}\Big(s_W{}^2+\sum_{i=1}^{I} w_i(\bar{y}_{i.}-\hat{\mu})^2+g(\tau^2)\Big)\Big\}\quad . \tag{2.6}$$

It is also possible to integrate the joint density (2.6) with respect to σ_W^2 to give:

$$p(\sigma_W^2|\tau^2,\underline{y})$$

$$\alpha\ (\sigma_W^2)^{-(N-3+\lambda)/2-1}\exp\Big\{-\frac{1}{2\sigma_W^2}\Big(s_W^2+\sum w_i(\bar{y}_i-\hat{\mu})^2+g(\tau^2)\Big)\Big\}\quad , \tag{2.7}$$

and:

$$p(\tau^2|\underline{y})$$

$$\alpha\ \frac{f(\tau^2)[\Pi(1+n_i\tau^2)^{-1/2}][\sum w_i]^{-1/2}}{(s_W^2+\sum w_i(\bar{y}_{i.}-\hat{\mu})^2+g(\tau^2))^{(N-3+\lambda)/2}}\quad . \tag{2.8}$$

This conditional posterior distribution (2.7) and the marginal posterior distribution (2.8) will be useful in computing posterior means and modes.

3. Choice of a prior distribution

Of course, $p(\mu,\sigma_B^2,\sigma_W^2)$ can be chosen subjectively to represent prior beliefs about the situation under consideration. It is demonstrated in section 4 that the convenient and tractable improper prior distributions considered here, which correspond to vague prior beliefs, do lead to estimators with desirable sampling properties.

There is no single convenient reference prior distribution for this model. For balanced data, $n_i = n$, Box and Tiao (1973) and others use Jeffreys diffuse prior. Stone and Springer (1965) criticize the Jeffreys prior and suggest an alternative. Neither can be used directly for unbalanced data as they depend on the common sample size n.

We chose to consider a non-informative prior distribution which seems appealing for unbalanced data. The model can be reparameterized into one depending on the mean μ, a single unknown variance σ_W^2 and the intraclass correlation coefficient $\rho = \sigma_B^2(\sigma_W^2 + \sigma_B^2)^{-1}$. The correlation ρ is the correlation between observations in the same group. A natural choice of prior distribution for σ_W^2 and ρ might be to take $\log\sigma_W^2$ uniform on $\sigma_W^2 > 0$ and ρ uniform on $0 < \rho < 1$ independently of σ_W^2. This leads to

$$p(\sigma_B^2, \sigma_W^2) = (\sigma_W^2 + \sigma_B^2)^{-2} \qquad \sigma_W^2, \sigma_B^2 > 0 . \tag{3.1}$$

This prior distribution will always lead to a proper posterior distribution.

We will also consider some informative prior distributions for σ_B^2 and σ_W^2. The choice of an informative prior distribution is clearly possible, although probably difficult in practice. In section 5.2 the prior distribution suggested in Hill (1965) is used, that is:

$$p(\sigma_B^2, \sigma_W^2) \propto (\sigma_W^2)^{-1} (\sigma_B^2)^{-(1+\alpha)} \exp\left(\frac{1}{\beta\sigma_B^2}\right) \qquad \sigma_W^2, \sigma_B^2 > 0. \tag{3.2}$$

The variances σ_B^2 and σ_W^2 are independent with σ_W^2 having an improper prior distribution and σ_B^2 having proper inverse gamma distribution with α and β chosen subjectively.

Lindley (1971) suggests a prior distribution with both σ_W^2 and σ_B^2 having inverse gamma distributions with the hyperparameters chosen subjectively. We could also, of course, choose an informative prior distribution for ρ and σ_W^2. We chose to restrict this study to a few limited choices of (3.2) with, $p(\mu)$ constant and μ independent of σ_B^2 and σ_W^2 in the prior distribution.

4. Computational Details

In this section, the approach to computing posterior modes and means is outlined. Techniques involving numerical maximization and integration in only one dimension are used.

The posterior mode of the joint posterior distribution $p(\sigma_B^2, \sigma_W^2 | \underline{y})$ was calculated. We denote the joint mode to occur at $(\hat{\sigma}_B^2, \hat{\sigma}_W^2)$. Define $\hat{\tau}^2 = \hat{\sigma}_B^2/\hat{\sigma}_W^2$ and w_i and $\hat{\mu}$ to be as in (2.3) with σ_B^2/σ_W^2 equal to $\hat{\tau}^2$. By inspection of (2.5) we have

$$\hat{\sigma}_W^2 = \frac{[s_W^2 + \sum w_i(\bar{y}_{i.} - \hat{\mu})^2 + g(\hat{\tau})]}{N + 1 + \lambda} \quad . \tag{4.1}$$

So if the above expression is substituted into (2.4), we can find $\hat{\tau}^2$ by maximizing the function (2.4) as a function of $\hat{\tau}^2$. Then $\hat{\sigma}_W^2$ is given by (4.1) and $\hat{\sigma}_B^2 = \hat{\tau}^2\hat{\sigma}_W^2$.

The maximum likelihood estimators can be computed in a similar way to the posterior modes. By inspection of the likelihood we see that $\hat{\mu}$ and $\hat{\sigma}_W^2$ can be expressed as functions of $\hat{\tau}^2$, the maximum likelihood estimator of τ^2. Therefore we only need to perform a one dimensional maximization.

The one dimensional method used to find posterior modes and MLE's was the routine 2XGSN from IMSL which is a golden section search method. This method works well in general as the functions to be maximized are typically unimodal

(Hill (1965)). On the rare occasions when the algorithm detected non-unimodality the function was maximized by evaluation on a grid.

Posterior means of σ_B^2 and σ_W^2 were also computed. They were computed via the relationships

$$E(\sigma_W^2 \mid \underset{\sim}{y}) = E_{\tau^2 \mid \underset{\sim}{y}} E(\sigma_W^2 \mid \tau^2, \underset{\sim}{y})$$

$$E({\sigma_B}^2 \mid \underset{\sim}{y}) = E_{\tau^2 \mid \underset{\sim}{y}} \tau^2 E(\sigma_W^2 \mid \tau^2, \underset{\sim}{y}) \quad .$$

The integration over τ^2 must be performed numerically. Posterior means were included in some of the simulations. As might be expected from the results of Klotz, Milton and Zacks (1969), who simulated data for the balanced case, the posterior means have, in general, very large mean squared error compared to, say, the ANOVA estimators. Results for posterior means are therefore not reported here.

The normal errors were simulated by the subroutine NORMAL on the University of Minnesota CYBER 74. One thousand replications of each experiment were run to calculate the mean squared errors of the estimators. On each data set the ANOVA estimators, the MLE's and the posterior modes were calculated.

5. Simulation Results

5.1 Non-informative prior distributions

The design of the simulation study in Swallow and Monanhan (1984) was used. That is, ten different patterns of values of the n_i's were used and seven different values of σ_B^2/σ_W^2 between 0 and 5.0. The patterns are a subset of a set of patterns designed and motivated in Swallow and Searle (1978).

For each simulated data set the ANOVA estimates and the posterior modes of $\rho(\sigma_B^2, \sigma_W^2|\underline{y})$ were calculated under each of the three non-informative prior distributions (5.1), (5.2) and (5.3) given below which are defined for σ_W^2, $\sigma_B^2 \geq 0$:

$$p(\sigma_B^2 , \sigma_W^2) \alpha \ (\sigma_W^2 + \sigma_B^2)^{-2} \tag{5.1}$$

$$p(\sigma_B^2 , \sigma_W^2) \alpha \ (\sigma_W^2)^{-1}(\sigma_W^2 + \sigma_B^2)^{-1} \tag{5.2}$$

$$p(\sigma_B^2 , \sigma_W^2) \alpha \ (\sigma_W^2)^{-1/2}(\sigma_W^2 + \sigma_B^2)^{-3/2} . \tag{5.3}$$

The distribution on μ is independent of σ_W^2 and σ_B^2 and is the improper prior $p(\mu)$ uniform. As discussed in Section 3, the prior distribution (5.1) has an interpretation in terms of σ_W^2 and the intra-class correlation. The distribution (5.2) is Jeffreys diffuse prior for a balanced experiment with n=1 and (5.3) is the alternative suggested in Stone and Springer (1965), again with n=1.

The three prior distributions are very similar, and the simulation results were, not suprisingly, very similar. The average bias and average squared error were almost identical. Only results for (5.1) are reported here. The average bias is given in Table 1 for the ANOVA estimators, the MLE's and the posterior modes. The ANOVA estimator of σ_B^2 used was the estimate obtained by equating mean squares and rounding up to zero if the unbiased estimate is negative. Table 2 gives the average squared error of each of the three different estimators for each of the two variances. The ratio of the squared error of the posterior mode to both the ANOVA estimator and the MLE in also given.

We first look at the estimation of σ_B^2. Inspection of Table 2 indicates that there seems to be a dramatic improvement in squared error using a posterior mode to estimate σ_B^2 instead of the ANOVA estimator. The improvement is as much as an

80% decrease in squared error for the Bayes mode over the ANOVA estimator. Even the smallest improvement was a 25% decrease. The improvement was greatest in patterns with either a small value of I, (P1, P2, P12 and P13) or a number of groups with $n_i = 1$. Patterns with I small can not be expected to give much information on σ_B^2, the between group variation. In patterns with groups with $n_i = 1$, information about σ_B^2 is more confounded with information about σ_W^2. So it would appear that the advantage of using a posterior mode of σ_B^2 as an estimator is greatest when the data provides little information about σ_B^2.

Swallow and Monahan reported that the estimator with the smallest MSE is generally the MLE. The posterior mode was found here to often have a smaller MSE than the MLE. Again the most noticeable improvement is for patterns with smaller values of I. The improvement seems greatest when σ_B^2 is small and the pattern is very unbalanced. In only 3 out of the 70 combinations of patterns and parameter values did the MLE have a smaller average squared error than the posterior mode.

Figure 1 is a graphical representation of the ratio of the squared errors for patterns P4, P5 and P7. The three patterns have the same total number of observations N, but range from almost balanced (P4) to very unbalanced (P7). The solid lines represent the ratio of the squared error of posterior mode of σ_B^2 to that of the MLE, and the dotted lines the ratio of the squared error of the posterior mode to that of the ANOVA estimator. These ratios are plotted against values of the intra-class correlation ρ. Both comparisons indicate an improvement by using the posterior mode and the improvement is greatest in the most unbalanced pattern, P7.

Table 1 is a table of the average biases. It is seen that the posterior modes do generally underestimate the variances on average. The bias in the estimation of σ_B^2 can be large, especially when σ_B^2 is large. These large biases, however, occur when the squared error of the ANOVA estimator of σ_B^2 is large, and despite the large bias in the posterior mode the squared error is smaller than that of the ANOVA estimator. The MLE's also have negative average biases in general. The bias in the posterior mode being more negative than that in the MLE is consistent with the prior weighting of the likelihood.

We now look at estimators of σ_W^2. Table 2 indicates that the mode of σ_W^2 is an improvement over the ANOVA estimator in terms of squared error. The decrease in squared error is around 10% in general, which is much less substantial than the decrease in squared error for σ_B^2. For the patterns in this study the degree of freedom within groups (N - I) is greater than the degrees of freedom between groups (I - 1) and so it could be expected that σ_W^2 would be estimated with more precision. In terms of squared error, the MLE's are slightly better than the posterior mode for values of σ_B^2 less than about 0.5. Posterior modes have smaller squared errors for larger values of σ_B^2. The biases in the MLE and posterior mode are small. The bias in the posterior mode is greatest where the improvement in squared error in greatest, that is for large values of σ_B^2.

We note that for very unbalanced patterns (eg P11 and P7) Swallow and Monahan report omitting up to 10% of the data sets from the experiment when convergence problems were encountered. No data sets were dropped from this study as a different optimization method was used which avoided convergence problems. Apart from this the results are in approximate accord.

5.2 Informative Prior Distributions

We now look at a very small simulation study of the properties of posterior modes under more informative prior distribution. The use of prior information is not uncommon in the estimation of variance components as both MINQUE and MIVQUE estimators require a prior specification of the unknown variance ratios.

The prior distributions we use here are not very informative and indeed do not have finite means. We use an inverse gamma informative prior distribution for σ_B^2 and we retain the non-informative prior for μ and σ_W^2. Thus $p(\mu, \sigma_B^2, \sigma_W^2)$ is given by equation (3.2). The three choices for (α, β), the hyperparameters of the inverse gamma distribution, in the simulation experiment were (1, 5), (1, 0.5) and (1, 0.1). Thus, the mean and variance are infinite but the prior mode is at 0.1, 1.0 and 5.0, respectively.

The value of σ_W^2 was kept at 1.0 and σ_B^2 was set at 0.1, 0.5, 1.0, 2.0 and 5.0. For each pattern, we have a 3 x 5 factorial structure for the combination of prior distribution and true value of σ_B^2. The simulations were again performed with 1000 replications but this time the same random deviates were used for each run of the same pattern. The simulations were done for only P1, P5, P7, P11 and P12. These patterns represent the most balanced (P1) and the most unbalanced (P7 and P11) of the patterns. Table 3 gives the average squared error for five estimates of σ_B^2. The five estimators are the ANOVA estimator, the Bayes posterior mode with the non-informative prior (5.1) and the Bayes posterior mode with the three informative priors with modes at 0.1, 1.0 and 5.0, respectively. Table 4 gives a similar table for the five estimators of σ_W^2. The MLE's were not included in this experiment.

The information in Table 3 indicates that the specification of a prior mode in this way can reduce the squared error in the estimation of σ_B^2 considerably. However, a prior mode that is very different than the true value can give larger average squared error than a non-informative prior. Similarly, we see from Table 4 that the posterior mode of σ_W^2 is also sensitive to the prior distribution on σ_B^2. The sensitivity is not as great as that of σ_B^2, and the improvement in using an accurate prior distribution is not as great.

In summary, this small simulation indicates that the use of a little prior information may lead to useful estimators. The posterior mode from an informative prior distribution can substantially dominate the posterior mode from a non-informative prior distribution, which in turn dominates the ANOVA estimators. If an experimenter is confident that σ_B^2 is close to zero, then using that information should improve the posterior mode as an estimator. Similarly, an experimenter can use prior information that σ_B^2 is large. If there is no such prior information then the posterior modes from the non-informative prior distribution (5.1) appear to have reasonable properties.

A more extensive study of the use of informative prior distributions might be useful. The problem is complicated as the properties of the posterior distribution depends on the values of both σ_W^2 and σ_B^2, not just their ratio σ_B^2/σ_W^2. The properties of the ANOVA estimators and the estimators based on a non-informative prior distribution depend only on the ratio.

A Bayesian analysis allows the use of prior information in a formal way rather than the MIVQUE and MINQUE methods, where prior information is used in a non-Bayesian way. The Bayesian approach allows formally for the specification of prior precision whereas MIVQUE and MINQUE do not. We may note that MIVQUE and MINQUE estimators may also be sensitive to a misguided prior specification. Swallow and Monahan noted that MIVQUE estimators with a default prior of zero for σ_B^2/σ_W^2 have terrible properties.

5.3 Point and interval estimation of $\tau^2 = \sigma_B^2/\sigma_W^2$

The ratio of the two variance components is often a parameter of interest and animal scientists refer to these ratios as heritabilities. We now look at a small simulation study of estimators of $\tau^2 = \sigma_B^2/\sigma_W^2$. We also consider interval estimators.

A natural interval estimate from a Bayesian viewpoint is a highest posterior density (hpd) region. These are the shortest regions of a given posterior density. They will always include the posterior mode. Confidence intervals for τ^2 can be constructed using the method given in Wald (1940). These confidence intervals have the specified content but can be empty on occasions where the mean square between groups is much smaller than the mean square within groups.

We note that there is no reason why hpd regions should have the properties of confidence intervals. Indeed, the simulation studies demonstrate that they do not.

The simulation study is a very limited investigation of the sampling properties of interval estimators. We take the extremes of a very unbalanced pattern (P11), a slightly unbalanced pattern (P8) and a totally balanced pattern (PB) with $n_i = 5$. In each case $I = 9$ and $\Sigma n_i = 45$. Values of τ^2 are taken to be 0.1, 0.5, 1.0 and 5.0. The content of the 90% hpd region is calculated as is the variability of the length. We also look at 90% confidence intervals for τ^2.

As computing the hpd regions is computationally intensive, only 500 data sets were used for each run. The same set of random deviates were used for each run of the same pattern. The calculations in this section were performed on a VAX 11/750 using double precision and routines rnor, dqagi and dqags from the core math library of the National Bureau of Standards (cmlib).

Table 5 is a summary of some of the simulation results for point estimation of τ^2. The average squared error and bias of the MLE and posterior mode (PM) under the prior distribution (5.1) are given. We see that in terms of mean squared error the posterior mode is again noticeably superior to the MLE. The negative bias in the posterior mode can be considerable however and can be considerably greater than that of the MLE. Further examination of the estimates shows that in each case there is almost perfect correlation between the MLE of

τ^2 and the posterior mode, the MLE being generally the larger of the two. The distribution of the MLE is seen empirically to have a long right tail and the effect of using the posterior mode is to shorten this tail considerably.

We now consider the sampling properities of confidence intervals and hpd regions for τ^2. For each data set a 90% confidence interval and an hpd region of probability 0.9 were calculated. The length of the intervals and whether or not the true value was included were calculated and some summary values are given in Table 6.

The coverage of an hpd region does not correspond to its posterior probability. For these configurations we see that if $\tau^2 = 5.0$, the coverage probability is around 80%. This is generally related to the negative bias which leads to the underestimation of τ^2. Alternatively, for τ^2 small, for example $\tau^2 = 0.1$, the coverage of an hpd region is close to 1.0. The hpd regions are much smaller, on average, than the confidence intervals.

For τ^2 small there is an appreciable probability that the confidence interval is empty. For example, in pattern 8, with $\tau^2 = 0.1$, this happened in 3% of the 500 simulations, for $\tau^2 = 0.5$ in 2% and in $\tau^2 = 1$ it did not happen at all. The hpd regions are clearly never empty.

6. Summary

The mode of the joint marginal posterior distribution of (σ_B^2, σ_W^2) from the non-informative prior distribution (5.1) is a viable estimator with desirable sampling properties. These estimates of σ_B^2 and σ_W^2 are as easy to compute as the maximum likelihood estimators and appear to have smaller squared errors in general, at the expense of a larger bias. These Bayesian estimators are generally much more efficient than the ANOVA estimators, although, of course, the ANOVA estimates can be easily computed by hand.

Prior information can be used in informative prior distributions and can improve the properties of the estimator. A Bayesian analysis allows the use for prior information in a formal way rather than the MIVQUE and MINQUE methods, where prior information is used in a non-Bayesian way. A Bayesian analysis also allows for the specification of prior precision. The posterior modes do seem to be very sensitive to the prior distribution even though the informative priors used were quite vague in that the mean and variance were infinite.

We also see that the Bayesian approach can lead to useful estimators of τ^2. Whilst hpd regions do not have the coverage probabilities of confidence intervals, they do lead to sensible interval estimates which are never empty.

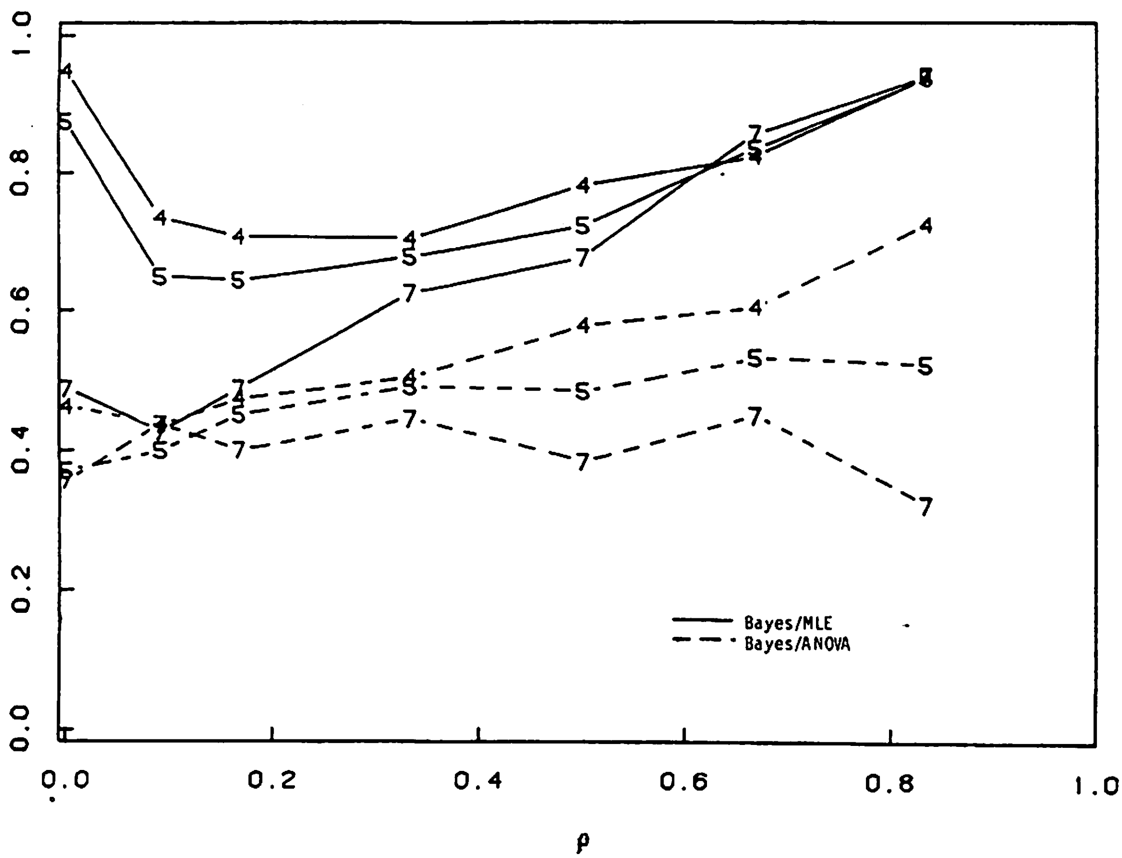

Figure 1: Ratio of the average squared errors of σ_B^2 plotted against $\rho = \sigma_B^2(\sigma_B^2 + \sigma_W^2)^{-1}$ for patterns P4 = (3,3,5,5,7,7), P5 = (1,1,5,5,9,9) and P7 = (1,1,1,1,13,13).

Table 1 Average biases of estimators of σ_W^2 and σ_B^2

	σ_B^2 ($\sigma_W^2=1$)	Estimators of σ_W^2 ANOVA	MLE	Bayes	Estimators of σ_B^2 ANOVA	MLE	Bayes
P1=(3,5,7)	0.0	.012	-.089	-.246	.084	.035	.035
	.1	.006	-.076	-.235	.062	-.021	-.031
	.2	-.010	-.074	-.234	.035	-.081	-.100
	.5	-.001	-.037	-.198	.053	-.188	-.270
	1.0	-.009	-.031	-.180	.035	-.354	-.573
	2.0	.023	.008	-.132	-.096	-.803	-1.261
	5.0	-.002	-.009	-.115	-.169	-1.820	-3.204
P2=(1,5,9)	0.0	.035	-.066	-.225	.098	.037	.033
	.1	-.001	-.079	-.237	.068	-.028	-.041
	.2	-.003	-.066	-.225	.074	-.069	-.102
	.5	-.003	-.038	-.196	.048	-.206	-.304
	1.0	.012	-.002	-.159	.046	-.430	-.643
	2.0	.007	-.007	-.145	-.147	-.917	-1.376
	5.0	-.001	-.004	-.106	-.050	-1.819	-3.309
P4=(3,3,5,5,7,7)	0.0	.017	-.051	-.143	.054	.032	.036
	.1	-.006	-.048	-.141	.033	-.009	-.010
	.2	.006	-.022	-.117	.014	-.048	-.057
	.5	-.012	.018	-.096	.024	-.020	-.107
	1.0	-.003	-.004	-.081	.004	-.197	-.361
	2.0	.014	.014	-.048	.020	-.361	-.766
	4.0	-.008	-.008	-.047	-.146	-.961	-2.125
P5=(1,1,5,5,9,9)	0.0	-.001	-.067	-.156	.055	.027	.031
	.1	.000	-.033	-.128	.037	-.016	-.016
	.2	-.008	-.035	-.128	.019	-.051	-.063
	.5	.006	-.005	-.093	.002	-.132	-.197
	1.0	-.008	-.005	-.080	.037	-.200	-.392
	2.0	-.001	.001	-.059	-.006	-.421	-.857
	5.0	-.002	.001	-.033	.055	-.892	-2.163
P7=(1,1,1,1,13,13)	0.0	-.007	-.058	-.148	.073	.028	.030
	.1	-.015	-.048	-.139	.044	-.016	-.024
	.2	.005	-.024	-.116	.026	-.055	-.078
	.5	.008	-.002	-.089	-.003	-.172	-.252
	1.0	-.011	-.007	-.081	.035	-.248	-.475
	2.0	-.005	.005	-.051	-.100	-.538	-1.015
	5.0	-.008	.000	-.029	.048	-.962	-2.344
P8=(3,3,3,5,5,5,7,7,7)	0.0	-.005	-.057	-.121	.042	.028	.032
	.1	-.000	-.025	-.092	.019	-.011	-.009
	.2	-.006	-.018	-.085	-.002	-.044	-.047
	.5	.006	.005	-.056	.015	-.066	-.112
	1.0	-.004	-.005	-.055	-.015	-.155	-.279
	2.0	.000	.000	-.037	.024	-.222	-.550
	5.0	.005	.005	-.017	.003	.536	-1.538
P9=(1,1,1,5,5,5,9,9,9)	0.0	.007	-.044	-.109	.039	.023	.027
	.1	.008	-.017	-.085	.021	-.012	-.010
	.2	.015	.002	-.066	.006	-.040	-.046
	.5	-.003	-.007	-.067	-.011	-.099	-.148
	1.0	-.007	-.005	-.054	-.005	-.163	-.306
	2.0	-.001	.000	-.036	.010	-.263	-.623
	5.0	-.000	-.000	-.019	.064	-.488	-1.537
P11=(1,1,1,1,1,1,1,19,19)	0.0	-.002	-.038	-.102	.059	.023	.026
	.1	-.002	-.022	-.088	.028	-.023	-.027
	.2	-.002	-.013	-.078	.010	-.059	-.076
	.5	.007	.003	-.054	.006	-.112	-.196
	1.0	.003	.010	-.037	.024	-.240	-.430
	2.0	-.004	.004	-.025	.068	-.313	-.776
	5.0	.000	.005	-.007	-.077	-.737	-1.890
P12=(2,10,18)	0.0	.001	-.046	-.137	.048	.016	.022
	.1	.011	-.020	-.116	.039	-.036	-.031
	.2	.023	-.001	-.098	.019	-.090	-.096
	.5	-.007	-.018	-.109	.026	-.204	-.268
	1.0	.001	-.005	-.088	-.001	-.387	-.576
	2.0	.001	.000	-.075	-.119	-.839	-1.261
	5.0	-.004	-.003	-.057	-.024	-1.700	-3.127
P13=(3,15,27)	0.0	-.020	-.048	-.112	.034	.012	.020
	.1	-.006	-.023	-.091	.015	-.046	-.036
	.2	-.017	-.027	-.093	.019	-.079	-.080
	.5	-.013	-.019	-.082	.015	-.194	-.250
	1.0	-.004	-.006	-.064	-.011	-.387	-.554
	2.0	-.002	-.001	-.050	.048	-.665	-1.134
	5.0	.001	.002	-.035	-.054	-1.772	-3.112

Table 2: Average Squared Error of estimators of σ_W^2 and σ_B^2

	σ_B^2 $(\sigma_W^2=1)$	Estimators of σ_W^2 ANOVA	ML	Bayes	Ratios $\frac{\text{Bayes}}{\text{ML}}$	$\frac{\text{Bayes}}{\text{ANOVA}}$	Estimators of σ_B^2 ANOVA	ML	Bayes	Ratios $\frac{\text{Bayes}}{\text{ML}}$	$\frac{\text{Bayes}}{\text{ANOVA}}$
P1=(3,5,7)	0.0	.170	.137	.148	1.081	.869	.038	.012	.007	.619	.197
	.1	.184	.145	.150	1.037	.818	.086	.030	.016	.534	.186
	.2	.160	.134	.141	1.054	.884	.128	.053	.032	.604	.252
	.5	.149	.141	.132	.936	.887	.519	.244	.154	.630	.296
	1.0	.167	.159	.140	.882	.840	1.554	.849	.567	.668	.365
	2.0	.181	.173	.138	.799	.764	4.906	2.670	2.208	.827	.450
	5.0	.174	.169	.140	.832	.808	25.486	14.873	13.460	.905	.528
P2=(1,5,9)	0.0	.182	.140	.143	1.024	.788	.050	.025	.009	.344	.171
	.1	.167	.140	.147	1.047	.876	.094	.038	.015	.411	.165
	.2	.164	.137	.139	1.016	.849	.201	.095	.040	.423	.199
	.5	.165	.152	.139	.914	.842	.574	.324	.174	.536	.303
	1.0	.182	.177	.143	.805	.785	2.147	.973	.638	.656	.297
	2.0	.175	.169	.138	.813	.786	5.555	2.836	2.439	.860	.439
	5.0	.187	.186	.149	.800	.796	36.160	16.095	14.223	.884	.393
P4=(3,3,5,5,7,7)	0.0	.089	.071	.076	1.078	.857	.012	.006	.006	.948	.463
	.1	.085	.072	.077	1.060	.902	.030	.018	.013	.734	.436
	.2	.091	.083	.079	.959	.873	.063	.042	.030	.706	.474
	.5	.079	.076	.072	.945	.907	.212	.152	.107	.703	.505
	1.0	.077	.077	.069	.896	.895	.581	.431	.336	.779	.578
	2.0	.087	.087	.075	.864	.867	2.130	1.565	1.287	.822	.604
	5.0	.078	.078	.071	.910	.910	11.078	8.542	8.008	.937	.723
P5=(1,1,5,5,9,9)	0.0	.089	.071	.079	1.110	.886	.013	.005	.005	.875	.370
	.1	.088	.075	.076	1.014	.863	.035	.022	.014	.648	.399
	.2	.081	.071	.072	1.019	.886	.063	.044	.028	.643	.451
	.5	.083	.079	.071	.907	.860	.256	.185	.126	.677	.491
	1.0	.075	.077	.068	.881	.902	.817	.548	.395	.721	.484
	2.0	.078	.078	.067	.862	.859	2.699	1.726	1.438	.833	.533
	5.0	.085	.089	.080	.900	.941	15.871	8.857	8.297	.937	.523
P7=(1,1,1,1,13,13)	0.0	.085	.069	.076	1.092	.887	.021	.016	.008	.489	.356
	.1	.086	.072	.075	1.049	.875	.043	.044	.019	.426	.437
	.2	.084	.072	.070	.970	.830	.096	.078	.038	.489	.399
	.5	.079	.075	.067	.889	.847	.342	.244	.152	.624	.445
	1.0	.078	.080	.070	.877	.903	1.320	.747	.505	.676	.383
	2.0	.085	.091	.077	.850	.911	3.933	2.067	1.768	.855	.450
	5.0	.085	.090	.081	.900	.953	29.311	10.041	9.443	.940	.322
P8=(3,3,3,5,5,5,7,7,7)	0.0	.059	.051	.056	1.103	.952	.007	.005	.005	.983	.630
	.1	.056	.050	.051	1.017	.916	.020	.014	.011	.822	.585
	.2	.050	.046	.047	1.001	.928	.033	.027	.022	.797	.646
	.5	.059	.058	.053	.904	.898	.140	.115	.087	.755	.620
	1.0	.057	.057	.052	.916	.917	.363	.294	.248	.845	.684
	2.0	.053	.053	.049	.921	.918	1.360	1.074	.915	.852	.673
	5.0	.058	.058	.054	.931	.931	8.094	6.296	5.784	.919	.714
P9=(1,1,1,5,5,5,7,7,7)	0.0	.058	.049	.052	1.079	.910	.007	.003	.003	1.001	.522
	.1	.057	.048	.048	1.003	.848	.021	.015	.012	.757	.560
	.2	.059	.052	.049	.934	.828	.040	.034	.025	.727	.624
	.5	.057	.053	.050	.931	.879	.150	.123	.093	.756	.618
	1.0	.054	.054	.050	.922	.931	.467	.349	.284	.815	.609
	2.0	.059	.059	.053	.905	.907	1.770	1.313	1.094	.833	.618
	5.0	.057	.057	.053	.930	.930	9.663	6.539	5.904	.903	.611
P11=(1,1,1,1,1,1,1,19,19)	0.0	.057	.048	.050	1.056	.881	.013	.011	.006	.528	.435
	.1	.057	.049	.049	.999	.865	.031	.033	.017	.519	.545
	.2	.057	.051	.049	.957	.858	.070	.067	.037	.550	.525
	.5	.057	.054	.047	.875	.838	.316	.227	.140	.616	.441
	1.0	.056	.056	.049	.872	.881	1.158	.562	.433	.771	.374
	2.0	.053	.057	.050	.887	.953	4.595	1.588	1.382	.870	.301
	5.0	.057	.061	.057	.938	.997	21.424	7.249	7.270	1.003	.339
P12=(2,10,18)	0.0	.072	.065	.071	1.078	.978	.013	.005	.004	.799	.283
	.1	.075	.069	.069	.996	.910	.053	.021	.013	.608	.243
	.2	.081	.074	.069	.937	.861	.112	.053	.032	.611	.288
	.5	.082	.079	.075	.953	.918	.499	.242	.151	.625	.303
	1.0	.070	.070	.064	.919	.910	1.630	.773	.547	.708	.335
	2.0	.072	.072	.065	.900	.907	5.758	2.525	2.155	.853	.374
	5.0	.072	.073	.066	.904	.917	35.946	15.282	13.240	.866	.368
P13=(3,15,27)	0.0	.044	.043	.048	1.116	1.084	.006	.002	.003	1.251	.474
	.1	.047	.046	.047	1.029	1.005	.032	.015	.010	.721	.323
	.2	.043	.042	.044	1.055	1.033	.096	.046	.030	.650	.310
	.5	.045	.046	.045	.992	1.002	.455	.228	.144	.634	.318
	1.0	.052	.052	.049	.947	.948	1.531	.691	.512	.741	.334
	2.0	.050	.049	.046	.936	.930	5.950	2.497	1.940	.777	.326
	5.0	.049	.058	.046	.933	.930	39.146	15.381	13.068	.850	.334

Table 3: Average Squared Errors of Estimators of σ_B^2 using informative prior distributions

	σ_B^2 ($\sigma_W^2=1$)	ANOVA	Bayes with (5.1)	Bayes estimator with prior mode at 0.1	1.0	5.0
P1=(3,5,7)	.1	.080	.015	.003	.610	14.293
	.5	.473	.152	.129	.283	12.413
	1.0	1.405	.545	.605	.159	10.325
	2.0	4.738	2.086	2.436	.800	6.999
	5.0	26.560	12.954	14.366	9.396	3.629
P5-(1,1,5,5,7,7)	.1	.036	.013	.002	.381	7.094
	.5	.267	.122	.112	.185	6.079
	1.0	.829	.404	.491	.166	5.012
	2.0	2.841	1.455	1.833	.774	3.504
	5.0	15.983	8.529	9.741	6.954	3.517
P7=(1,1,1,1,13,13)	.1	. 44	.019	.004	.483	8.042
	.5	.295	.148	.127	.232	6.895
	1.0	.966	.484	.580	.169	5.676
	2.0	3.488	1.680	2.214	.757	3.906
	5.0	20.562	9.166	10.990	7.086	3.345
P11=(1,1,1,1,1,1,1,19,19)	.1	.032	.019	.003	.394	5.527
	.5	.294	.133	.120	.194	4.682
	1.0	1.018	.416	.513	.165	3.818
	2.0	3.777	1.359	1.812	.693	2.676
	5.0	22.591	6.767	8.144	5.600	3.103
P12=(2,10,18)	.1	.054	.015	.002	.544	13.015
	.5	.469	.150	.122	.240	11.233
	1.0	1.592	.534	.578	.140	9.250
	2.0	5.867	2.036	2.392	.832	6.122
	5.0	35.071	12.735	14.337	9.656	3.405

Table 4: Average Squared Errors of Estimators of σ_W^2 using informative prior distributions

	σ_B^2 ($\sigma_W^2=1$)	ANOVA	Bayes with (5.1)	Bayes estimator with prior mode at 0.1	1.0	5.0
P1=(3,5,7)	.1	.166	.140	.145	.142	.148
	.5	.166	.134	.184	.146	.148
	1.0	.166	.132	.233	.149	.149
	2.0	.166	.131	.251	.152	.149
	5.0	..166	.134	.254	.156	.151
P5=(1,1,5,5,7,7)	.1	.079	.072	.072	.071	.073
	.5	.079	.069	.084	.072	.073
	1.0	.079	.069	.095	.074	.074
	2.0	.079	.070	.103	.077	.074
	5.0	.079	.072	.099	.079	.076
P7=(1,1,1,1,13,13)	.1	.078	.069	.068	.069	.071
	.5	.078	.069	.081	.071	.071
	1.0	.078	.069	.099	.073	.072
	2.0	.078	.071	.126	.076	.073
	5.0	.078	.075	.131	.080	.075
P11=(1,1,1,1,1,1,1,19,19)	.1	.057	.049	.049	.051	.052
	.5	.057	.049	.057	.052	.052
	1.0	.057	.053	.067	.053	.053
	2.0	.057	.054	.082	.056	.053
	5.0	.057	.056	.084	.059	.055
P12=(2,10,18)	.1	.070	.069	.068	.067	.068
	.5	.070	.066	.075	.068	.068
	1.0	.070	.065	.079	.069	.068
	2.0	.070	.065	.082	.069	.069
	5.0	.070	.065	.078	.070	.069

Table 5: Average bias and squared error of the maximum likelihood (ML) and posterior mode (PM) estimator of $\tau^2=\sigma_W^2/\sigma_W^2$

	σ_B^2 ($\sigma_W^2=1$)	Bias ML	Bias PM	Squared Error ML	Squared Error PM
P8=(5,5,5,5,5,5,5,5,5)	0.1	-.00	-.02	0.02	0.01
	0.5	-.06	-.14	0.14	0.10
	1.0	-.09	-.39	0.41	0.30
	5.0	-.37	-1.70	7.71	6.38
P8=(3,3,3,5,5,5,7,7,7)	0.1	-.01	-.02	0.02	0.01
	0.5	-.06	-.14	0.13	0.10
	1.0	-.10	-.30	0.39	0.30
	5.0	-.37	-1.72	7.05	6.13
P11=(1,1,1,1,1,1,1,19,19)	0.1	-.01	-.03	0.05	0.02
	0.5	-.10	-.24	0.29	0.15
	1.0	-.14	-.46	0.76	0.48
	5.0	-.32	-2.00	10.43	8.49

Table 6: Content and average length of 90% confidence intervals (CI) and 90% highest posterior density regions (HPD) for τ^2

	σ_B^2 ($\sigma_W^2=1$)	Content CI	Content HPD	Average Length CI	Average Length HPD
P8=(5,5,5,5,5,5,5,5,5)	0.1	.92	.99	0.7	0.6
	0.5	.91	.94	1.9	1.3
	1.0	.90	.89	3.2	2.1
	5.0	.90	.82	14.2	8.0
P8=(3,3,3,5,5,5,7,7,7)	0.1	.91	1.00	0.8	0.6
	0.5	.90	.94	1.9	1.3
	1.0	.91	.88	3.3	2.1
	5.0	.90	.80	14.3	8.0
P11=(1,1,1,1,1,1,1,19,19)	0.1	.90	1.00	2.0	1.1
	0.5	.92	.99	3.2	1.8
	1.0	.91	.93	4.7	2.6
	5.0	.90	.81	16.1	8.6

REFERENCES

Box, G.E.P and Tiao, G.C. (1973), Bayesian Inferences in Statistical Analysis, Addison-Wesley, Reading, Mass.

Hill, B.M. (1965), Inference about variance components in the one-way model, Jour. Amer. Statist. Assoc., 60, 806-825.

Hill, B.M. (1977), Exact and Approximate Bayesian Solutions for inference about variance components and multivariate inadmissibility, in New Developments in the Applications of Bayesian Methods, ed. A. Ayka and C. Brumat, North-Holland, 129-152.

Klotz, J.H., Milton, R.C. and Zacks, S. (1969), Mean Square Efficiency of estimators of Variance components, Jour. Amer. Statist. Assoc., 64, 1383-1402.

Lindley, D.V. and Smith, A.F.M. (1972), Bayes estimates for the linear model, Jour. Roy. Statist. Soc. Ser. B., 34, 1-41.

Searle, S.R. (1979), Notes on Variance Component Estimation: A detailed account of maximum likelihood and kindred methodology, Paper BU-673-M, Biometrics unit, Cornell University.

Skene, A.M. (1983), Computing Marginal Distributions for the Dispersion Parameters of Analysis of Variance Models, The Statistician, 32, 99-108.

Stone, M. and Springer, B.G.F. (1965), A paradox involving quasi-prior distributions, Biometrika, 52, 623-627.

Swallow, W.H. and Monahan, J.F. (1984), Monte Carlo comparison of ANOVA, MIVQUE, REML and ML Estimators of Variance components, Technometrics, 26, 47-57.

Swallow, W.H. and Searle, S.R. (1978), Minimum Variance Quadratic Unbiased Estimation (MIVQUE) of Variance Components, Technometrics, 20, 265-272.

Tiao, G.C. and Tan, W.Y. (1965), Bayesian analysis of random-effect models in the analysis of variance I. Posterior distribution of variance components, Biometrika, 52, 37-53.

Wald, A. (1940), A note on the analysis of variance with unequal class frequencies, Annals of Math. Statist., 11, 96-100.